

METHOD AND APPARATUS FOR ENCODING AND DECODING
DIGITAL REPRESENTATIONS OF WORKS

FIELD OF INVENTION

5 The present invention relates generally to the field of
protecting copyrighted digital works from piracy and more
specifically to a method and an apparatus for rendering
works unique in such a way that each and any digital
representation treated in accordance with the invention
10 may be identified as to its source, legitimate recipient,
or any other information vital to the rights protection
of the digital work in question.

BACKGROUND

Digital representations, of which musical works or video
15 information are examples, commonly rely on a number of
known algorithms and methods which compress the infor-
mation content in order to reduce the need for digital
storage space or transmission bandwidth. Such repre-
sentations are commonly converted from analogue to
20 digital (binary) form by a method of sampling the
original wave forms at a rate which is sufficiently
higher than the wave forms being imaged. Simplified the
techniques being used are based on the same principle as
when statistical points are plotted on a chart and are
25 later bound together forming a curve representing a
trend. Waves of audio or video information which are very
closely related in time, such as adjacent sampling points
are considered to be neighbors in amplitude, being part
of a trend going up or going down. This way the sampling
30 frequency itself determines the amount of accuracy or
fidelity compared to the unbroken analogue wave form. A
digital sampling thus has a built in potential loss of

information which may or may not be detectable when the information later is retrieved. A state of transparency is aimed at, which ideally means that a perfect copy of the original analogue data should be possible to retrieve
5 after digital encoding and decoding.

To give further examples, the sampling frequency being used in Compact Disc recordings is 44.1 kHz. Roughly that gives a potential of retrieving 20 kHz per channel in a stereo recording which is an approximate upper limit of
10 what a person with perfect hearing can perceive.

Digitally distributing music at this rate with 16 bits per sample requires a bandwidth of almost 1.5 megabits per second and it is therefore often desirable and needed to reduce the bandwidth requirement by further
15 compression. The currently most popular format is MP3 which uses a psychoacoustic model for its compression. The psychoacoustic part of the compression relies on algorithms that discard information which it is considered that the human ear cannot hear anyway and which
20 therefore in reality is extraneous information. An example is a tiny sound which occurs at or in the immediate proximity of a very strong sound. The human ear will have difficulty distinguishing the faint sound while being blocked or distracted by the much stronger sound,
25 such as hearing a needle fall while the burglar alarm siren is on to give an extreme example.

Similar techniques are applied in case of visual materials, in order to remove information that the human eye cannot adjust to such as weak reflections of surrounding
30 objects in front of a sunlit background.

After the step of removing such information, a further compression step is usually applied which may or may not consist of lossless compression such as the use of Huffman tables to avoid that repeating patterns in the
5 information consumes bandwidth unnecessarily.

In a number of existing patents protection is attempted to be done before as well as after the distribution of materials. Methods of protection before distribution may involve encryption of the material in such a way that
10 only persons who are in the possession of specific digital keys or special hardware equipment can access the material. Protection of copyrighted material after distribution may entail methods and equipment in order to coerce potential transgressors not to engage in illegal
15 copying or redistribution by introducing potential discovery.

One such method is described in U.S. Pat. No. 6,330,672 to David Hilton Shur issued in 2001, a method for watermarking digital bit streams. The referenced patent relies
20 on various principles of superimposing or integrating decipherable signals or marks which can later serve to identify the work in question or the individual behind illegal redistribution.

In any digital transmission, whether it concerns sound,
25 radio or other frequencies, there are a number of known parameters that can be varied in order to perceptibly or imperceptibly convey embedded messages. Frequency modulated radio transmissions are an example where the frequency of a carrier wave is altered in pace with
30 speech or music and which variation is later retrievable as the original message. Amplitude, phase and pulse are

other examples of parameters that can be so varied. The intended message can then be further concealed by various methods of encryption.

5 In the case of the referenced U.S. Pat. No. 6,330,672, a key may accompany the message as an indicative of a certain mark or marks which have been hidden inside the original message, a virtual watermark, thereby making it possible to reveal the source or identity of the digital material so marked.

10 A problem to be solved is thus to provide a method and an apparatus for encoding digital information which both provide an unique new original while making this encoding virtually impossible to find by a fraudulent person.

SUMMARY OF THE INVENTION

15 An object of the present invention is to provide a method and an apparatus for rendering works unique in such a way that each and any digital representation treated in accordance with the invention may be identified while making this identification virtually impossible to find
20 by a fraudulent person.

The invention is based on the realization that the hardest thing to find is something which is not there. Making use of this realization, the current invention processes any digital material in such a way to rendering
25 it unique as an original and imperceptibly encoding a message by way of removing objects or tiny fractions of objects, fractions of wave forms or bits from the original material. By removing should be understood actual removal in time of information, not merely
30 replacing existing information with other information.

According to the invention there is provided a method of encoding digital representations of works comprising the steps of providing a digital representation of a work, and removing pieces of information from said digital
5 representation in a predetermined pattern.

There is also provided a method of decoding digital representations of works comprising the following steps: synchronizing a customer original digital representation of a work with a true original digital representation of
10 a work, identifying a first piece of information missing in said customer original digital representation as compared with said true original digital representation, recording an identification of said first piece of information missing; and identifying and recording second and
15 further pieces of information missing in said customer original digital representation by repeating the identifying and recording steps.

There is also provided an apparatus for performing encoding of digital representations of works comprising:
20 means for providing a digital representation of a work, and means for removing pieces of information from said digital representation in a predetermined pattern.

There is also provided an apparatus for decoding digital representations of works comprising: means for
25 synchronizing a customer original digital representation of a work with a true original digital representation of a work, means for identifying a first, second and further pieces of information missing in said customer original digital representation as compared with said true
30 original digital representation, and means for recording

an identification of said first, second, and further pieces of information missing.

With the inventive methods and apparatus the above-mentioned drawbacks of prior art are eliminated or at least mitigated. The method provides for an identification of digital representations which is virtually impossible to locate and remove.

In a particularly preferred embodiment, the predetermined pattern is repeated, making removal thereof from an illegal copy virtually impossible.

Further preferred features are defined in the dependent claims.

BRIEF DESCRIPTION OF DRAWINGS

The invention is now described, by way of a non-limiting example, with reference to the accompanying drawings, in which:

FIG. 1 shows two wave forms that are identical but which have no coinciding points due to a slight offset of time on one of the wave forms.

FIG. 2 shows a number of random objects.

FIG. 3 shows a musical wave form where a tiny fraction has been cut out.

FIGs. 4A and 4B are flow charts describing the essential workflow of an exemplary embodiment of the invention.

FIG. 5 is a flow chart showing the encoding process according to the present invention.

FIG. 6 illustrates how a missing sampling is equalized by standard decoding techniques.

FIGS. 7A-D show an example of a simplified pattern and how it is matched while mining proper identification data
5 from an absence coded digital representation.

FIG. 8 illustrates a wave form where extraneous information has been added.

FIG. 9 shows the phasing of two wave forms.

FIG. 10 is a diagram explaining the absence coding
10 principles.

FIG. 11 is a flow chart describing the basic absence coding detection.

FIG. 12 is an overall diagram of an apparatus for performing the inventive method.

15 DETAILED DESCRIPTION OF THE INVENTION

In the following a detailed description of preferred embodiments of the present invention will be given. This invention presents an effective variant of how the digital representation of any copyrighted material or
20 works can be imperceptibly altered in order to render the representation unique and identifiable. In order to embed a message or identification into a digital representation, it is usually necessary to add or superimpose the desired information onto the original material. More
25 or less sophisticated methods can then be used in order to retrieve the information at a later stage. Using any such method, the original material would then be intact, whether encrypted or not, together with the additional

information. Someone who intends to steal the material would then need to restore the material to an unencrypted state as well as remove the identity information.

According to the principles of the present invention, the
5 most effective way of protecting digital works in this context is first of all to render the material to be distributed unique in such a way that no two "copies" are the same but all representations of a material are unique as virtual originals. In the following, the expression
10 "true original" will be used to denote the original that is encoded to create a "customer original", which is the information that is distributed to customers. All originals can be provided in any convenient form, such as a computer file.

15 Also, in the present description the word "work" will be used. Although a music piece will be used as an example, it will be appreciated that this word also encompasses other kinds of information, such as video recordings, text information etc. It should also be noted that digi-
20 tized work in this context refers to any work, whether originally in analogue or digital form, which is represented by a bit stream or sampling of the original material whether previously processed by other software or devices or not.

25 Using a music piece as an example to illustrate this first principle, this step can be accomplished for instance by altering the phase of the music. FIG. 1 shows two wave forms that are identical but which have no coinciding points due to a slight offset of time on one
30 of the wave forms. In the example, the difference between the forms is easily detectable and the illustration is

made only for the purpose of making the underlying principle understandable.

The second important principle of the present invention is the fact that, unless there is a useful reference, the
5 hardest thing to find is something which is not there. FIG. 2 shows a number of random objects. It is impossible, just by looking at FIG. 2 to determine if the number of objects is correct or if they are the right objects. Making use of these principles, the current
10 invention processes any digital material in such a way to render it unique as a customer original and imperceptibly encoding a message by way of removing objects or tiny fractions of objects, fractions of wave forms or bits from the original material.

15 The basic principles of the invention is further illustrated by FIG. 3 which shows a musical wave form where a tiny fraction has been cut out, which in FIG. 3 is equivalent to a missing sampling. Due to the standard techniques used in decoding, for example music encoded
20 with MP3 compression, the missing sampling can not be detected in the resulting music output.

FIG. 4A is a flow diagram which describes the essential encoding workflow of an exemplary embodiment of the invention. In step 10, a customer orders a digital work,
25 such as a music file, via online sales on e.g. the Internet. The customer has to register himself or herself, step 20. During the registration process information for identifying the customer must be given.

After the registration process, the true original is re-
30 trieved, step 30, and a copy wave file is made from the

true original, step 40. The provision of a customer original involves running the customer ID through a coding processor, step 50, whereby a series of algorithmic deletion patterns is obtained, step 60. In step 5 70, the deletion pattern process is applied to the customer wave file where after data padding compensating for deleted bytes is performed in step 80. The unique customer original which is a modified copy of the true original is then ready for downloading in step 90. When 10 the customer has downloaded his or her original in step 100, this file is optionally removed from the server in step 110

FIG. 4B is a flow diagram which describes the essential decoding workflow of an exemplary embodiment of the 15 invention. When a possibly illegal copy of a work has been found, a legal department investigating file sharing services is consulted in step 120. The content of the possibly illegal copy, such as a particular song, is identified in step 130. The possibly illegal copy is then 20 processed against the true original in step 140 and the signature patterns are found and decoded in step 150. A database matching is performed in step 160, wherein the ID of registered customers are used. When a match is found in step 170, the legal department confronts the 25 identified customer with evidence in step 180.

Essential for the workability of the present invention is that the device utilized for the absence encoding according to the invention has a processing module or processor which imperceptibly alters the true original in 30 such a way that each customer original so produced becomes a unique and identifiable copy or virtual new

original. Steps 30-80 of FIG. 4A are exemplified in FIG. 5, showing the encoding process.

FIG. 6 illustrates how a missing sampling is equalized by standard decoding techniques. One of the main purposes of the technique employed by the present innovation is thus to make it possible to imperceptibly code a message into any digital bit stream or any sampled wave form as distinct from hitherto utilized and patented techniques which utilize perceptual coding.

- 10 To the upper left in the figure there is shown the original sequential profile. The sequence values 900001 and 900004 are marked as two pieces of information that are to be removed in the encoding process. The encoding algorithm ensures that the removal of the information
- 15 does not degrade the quality of the encoded work in a perceptual way. Also, the algorithm may also be adapted not to remove a piece of information with identical value as an adjacent piece of information to ensure that decoding will be possible.
- 20 With sequence values 900001 and 900004 removed, the encoded profile will look as is shown to the lower right in FIG. 4. The encoded profile is shorter than the original sequence. This is corrected by means of padding, as will be described below with reference to FIG. 8.
- 25 The invention also considers the possibility that someone bent on illegal distribution may further alter the digital work in order to cut out or disguise the origin of the work or any inherent identity information. To prevent such occurrence the invention prescribes a
- 30 repetition of the absence encoding in such a way as to

constitute a pattern, exemplified in FIG. 7A, which is visually or audibly imperceptible to the end user during playback of the material. This pattern can take any suitable form, i.e., there can be one or several
5 successive deletions of samples, the distance between deleted samples can vary etc.

While it is recognized that recurring patterns is a lead in to the decoding of any coded message, the patterns introduced by the present invention is much less
10 vulnerable to detection as they consists of elements which are not there. Again, it is extremely hard to detect something which is not there unless one knows that it should be there. Through the use of patterns, the present invention makes it virtually impossible to cut
15 out or alter parts of the work being protected in order to destroy the identifiable originality of the material. The deletion thus ensures that the pattern can be found even though a fraudulent owner removes part of the information. As an example, the deletion pattern can be
20 introduced ten times in a wave file of a music work with a duration of three minutes.

The wave form of the customer original is shown in FIG. 7B. Padding to compensate for gaps is not shown. When expanding the customer original wave form by applying the
25 deletion pattern in FIG. 7A on this wave form, the wave form shown in FIG. 7C is obtained, showing the absence gaps. This can be compared with the true original wave form shown in FIG. 7D.

Although the number of distinct patterns used while
30 implementing the present invention is not limited and can consist of anything between one and millions of patterns,

it is suggested for the practicability of the implementation, that the number of patterns is kept manageable, perhaps in the order of a hundred. The patterns which preferably are distributed throughout the audio, video, or bit stream, serve several purposes. They prevent intentional distortion of the information content for the purposes of masking or destroying the inherent identity information while they also, together with a check sum figure, serve to validate the imperceptibly coded absence information.

In order to prevent that the described patterns open doors to alter the originality of a media file for example by comparing several different files in order to obtain information by mapping differences, wave form fragments or extraneous information may preferably be added in order to pad the length of a file, a segment of a file or a bit count. Any extraneous information so added and exemplified in FIG. 8 does not contain any information and relates to the present invention only as a means of disguising the positioning of elements belonging to the absence coding.

FIG. 9 shows the restoration of an absence encoded wave file shown in dashed lines by the insertion of a previously deleted sample. Thereby the absence encoded wave file will correspond to the original wave file shown in continuous lines.

The absence coding principles will be further explained in the following with reference to FIG. 10. The first signal absence is detected by synchronization and comparison with the true original. When an absence is

found the coded reference segment is retrieved by utilization of either of:

- the starting level of the absence, or
- the trailing level of the absence, or
- 5 - the starting time of the absence, or
- the trailing time of the absence, or
- the level of the omitted segment in the true original, or
- any combination of the above.

10 A decoding process according to the invention will now be described with reference to FIG. 11. In step 200, a pre-analysis is done of the files to be checked for identity. The purpose of the pre-analysis is to group e.g. music files into categories in order to lower the amount of
15 files that will have to be cross checked against each other. Files may be grouped according to length, as well as a number of other characteristics, such as beat.

In step 210, the comparison of a file against a master file, starts with a phasing of the files in step 220
20 against each other in order to determine a common starting point where the wave form of the file to be compared, closely matches a segment of the master file. This initial phasing is done by taking time as well as playback speed into consideration. The purpose of this
25 step is to enable identification also of files which have been distorted, whether on purpose or not.

In step 230, when a close match has been found between a segment of the master file and corresponding segment of the file to be compared (the compare file), envelop reconstruction is done in order to restore the compare
5 file to its original status and eliminate losses caused by prior compression in the distribution chain. Coded absences in the compare file are differentiated and are not restored at this point in time.

An absence code detection is performed in step 240. From
10 step 230 above, the compare file has been restored to be almost identical to the master file. Any difference is part of the potential absence coding of that file. When such an absence is detected, the detection process automatically goes on high alert and switches to pattern
15 detection. This signifies that subsequent absences must follow the initial absence according to one of a number of predetermined patterns. Each absence detected in the compare file is checked to find if it is part of one of the predetermined patterns or not.

20 When a pattern match has occurred the previous steps are looped throughout the file. When a pattern match occurs, the coded reference number is retrieved according to the basic principles shown in FIG. 10.

Step 250 is a verification step. The check sum of the
25 reference number retrieved in step 240 is calculated in order to determine that the reference number is valid. If it is found to be valid, a data base look-up is done in step 260 in order to find the identity corresponding to the reference number.

Provided that the reference number in step 250 checked out and a valid identity could be retrieved from the data base, an identity report is written out or is otherwise presented in step 270. In the case of comparison of files
5 suspected of copyright violations, the report, as in FIG. 11, is a violation report.

An overall diagram of an apparatus for performing the inventive method is shown in FIG. 12. A computer 300 having the hardware necessary for performing the required
10 calculations is used. The computer is preferably connected to a computer network, such as the Internet. The same network can be used by a number of users, one of which is represented by a computer 310 in the figure, for downloading works protected with the method according to
15 the invention.

A preferred embodiment of a method and an apparatus for encoding digital representations of works according to the invention have been described. The person skilled in the art realises that this could be varied within the
20 scope of the appended claims.

Digital sampling of a work has been described as part of the inventive method. It will be appreciated that both constant and variable bit rate sampling are feasible.

Downloading over the Internet has been described as a way
25 of distributing digital representations of works protected by absence encoding. It will be appreciated that digital representations of works distributed through other channels, such as music shops etc., can be protected by means of absence coding. In that case, each
30 digital representation distributed this way is uniquely

- coded and this unique code is noted at the time of selling the digital representation. For example, the customer is asked to identify himself or herself at the time of purchase and the identity is associated with the unique code in a register. Alternatively or additionally the identity of the work, such as the name of the record and/or the artist, could be encoded into the digital representation of the work so as to enable identification thereof if a manipulated illegal copy is found.
- 10 The present invention makes no claims as to the content of the actual message or data being coded. For practical purposes it may be convenient to adapt the original work being processed in accordance with the invention in such a way that the missing fragments indicate a specific
- 15 number with a check sum figure which points to a certain record in a central data base.